

Ein VBA-Projekt für Excel zur Unterstützung des Verständnisses des adaptiven Testens

Dr. Helmut Stauche und Anita Polet M.A.

Institut für Erziehungswissenschaft der Universität Jena

Gliederung:

1.	Theoretische Vorbemerkungen.....	2
1.1	Itemauswahl beim adaptiven Testen.....	3
1.1.1	Tailored testing.....	3
1.1.2	Branched testing.....	4
1.2	Vor- und Nachteile adaptiven Testens.....	5
2.	Das VBA-Projekt „Herstellung, Durchführung und Auswertung eines adaptiven Leistungstests“.....	7
2.1	Die Datei „adaptiv.xls“ als Gestaltungsrahmen.....	7
2.2	Der vorgefertigte Test „kopfrechnen_bvj_reha.xls“.....	9
2.2.1	Die Testitems.....	9
2.2.2	Die Testeichung.....	13
2.2.3	Der Testeinsatz.....	15
	Referenzen.....	17

1. Theoretische Vorbemerkungen

Der Grundgedanke des adaptiven Testens ist ein flexibler Testprozess, der durch das Antwortverhalten des Probanden mitgesteuert wird.

Nach Frey (2008, S. 262) wird adaptives Testen definiert als ein "...spezielles Vorgehen bei der Messung individueller Ausprägungen von Personenmerkmalen¹, bei dem sich die Auswahl der zur Bearbeitung vorgelegten Items am Antwortverhalten des untersuchten Probanden orientiert".

Frey veranschaulicht den Grundgedanken des adaptiven Testens am Beispiel einer mündlichen Prüfung. Ein Prüfer beginnt eine mündliche Prüfung in der Regel mit mittelschweren Fragen, der weitere Prüfungsverlauf wird durch die Antworten des zu Prüfenden mitbestimmt. Beantwortet er die mittelschweren Fragen richtig, wird der Prüfer im folgenden Verlauf der Prüfung schwerere Fragen stellen, um herauszufinden, wie tief und umfassend das Wissen des Prüflings ist. Kann der Prüfling jedoch bereits die mittelschweren Aufgaben nicht oder nur unzureichend beantworten, wird der Prüfer den Schwierigkeitsgrad seiner Fragen herabsetzen, um zu prüfen, ob das vorhandene Wissen des Prüflings zum Bestehen der Prüfung ausreicht oder diese wiederholt werden muss.

Diese Anpassung des Schwierigkeitsgrades der Fragen ist der Grundgedanke des adaptiven Testens. Im Gegensatz zu den meisten Testverfahren wird dem Probanden keine vorher festgelegte Menge an Items vorgelegt, die es vollständig zu bearbeiten gilt, sondern er² erhält lediglich solche Items, die möglichst viele Informationen bezüglich des zu messenden Merkmals liefern. Im oben beschriebenen Prüfungsbeispiel wird deutlich, dass adaptives Herangehen nicht nur im mittleren Ausprägungsbereich eines Leistungs- oder Persönlichkeitsmerkmals misst, sondern auch eine differenzierte Messung des zu untersuchenden Merkmals in den Randbereichen ermöglicht wird. Diese differenzierte Messung eines Merkmals ist natürlich auch mit einem Testverfahren möglich, bei dem vorher alle zu bearbeitenden Items festgelegt sind, nur müssten dann viel mehr Items vom Probanden bearbeitet werden.

¹ Der Terminus *Personenmerkmal* wird von Frey übergreifend benutzt. Klarer ist es, von Leistungsmerkmalen zu sprechen, denn ein adaptiver Test ist immer ein Leistungstest.

² Die weibliche Form ist immer mitgemeint.

1.1 Itemauswahl beim adaptiven Testen

Für die Itemauswahl aus dem vorbereiteten Itempool hält das adaptive Testen unterschiedliche Strategien bereit. Allgemein lässt sich sagen, dass zu den Testzeitpunkten Beginn und Ende sowie während der gesamten Testphase durch einen Algorithmus Regelungen vorgenommen werden.

Zu Beginn eines Testes liegen im Normalfall noch keine Informationen über das zu untersuchende Merkmal eines Probanden vor. In der Regel wird als Startitem ein Item mit mittlerer Schwierigkeit gewählt, um eine Überforderung und daraus möglicherweise resultierende Frustration zu vermeiden.³

Die Itemauswahl während des Testes folgt der Festlegung, dass der Proband immer dann ein leichteres Item zur Beantwortung erhält, wenn er das vorherige nicht richtig lösen konnte und ein schwereres Item, wenn er das vorherige richtig löste. Die Auswahl des nachfolgenden Items kann dabei nach einer der beiden Varianten des adaptiven Testens erfolgen: dem tailored testing und dem branched testing (Pospeschill 2010, S. 154f.). Beide sollen im Folgenden näher betrachtet werden.

1.1.1 Tailored testing

Das Verb „tailored“ lässt sich am besten mit „maßgeschneidert“ wiedergeben. Die Itemauswahlstrategie ist beim tailored testing nicht vorgegeben, sondern flexibel. Es folgt immer das Item auf die Beantwortung des vorherigen Items, von dem die maximale Information in Bezug auf das zu messende Leistungsmerkmal erwartet werden darf. Diese Itemauswahlstrategie ist somit abhängig von der jeweiligen Probandenreaktion. Anders als bei der „Maßschneiderei“ im Schneiderhandwerk läuft jedoch kein einmal festgesetzter – nach der Abnahme der Körpermaße des Kunden – Prozess ab sondern ein kontinuierlicher und mit jedem Item neu festgelegter Prozess.

Es leuchtet ein, dass ein solches Vorgehen computergestützt ablaufen muss, da mit jeder Itembeantwortung der Personenparameter geschätzt wird, um möglichst nah mit der Schätzung an den wahren Wert des Parameters zu kommen. Beim tailored testing kommen oft auch Testabbruchkriterien zum Einsatz, auf welche weiter unten näher eingegangen wird (Kubinger & Jäger 2003, S. 4f.).

³ Sind jedoch Vorinformationen über das zu untersuchende Merkmal des Probanden aus parallelen oder früheren Tests vorhanden, so können diese Informationen zu einer vorläufigen Schätzung der Merkmalsausprägung herangezogen werden und so ein passendes Item mit gewünschtem Schwierigkeitsniveau aus dem vorhanden Itempool gewählt werden (Pospeschill 2010, S. 154).

1.1.2 Branched testing

Das Verfahren des branched testing wird i.a. als weniger optimal angesehen. Hierbei werden Items zu unterschiedlichen Itemgruppen zusammengefasst, jeder Proband löst zunächst eine Itemgruppe. Je nachdem, wie viele Items dieser Gruppe vom Probanden richtig gelöst wurden, wird ihm in Abhängigkeit davon als nächstes eine leichtere, schwerere oder im Schwierigkeitsgrad gleiche Itemgruppe zur Beantwortung vorgegeben. Parameterschätzungen erfolgen vor dem Test und nicht wie beim tailored testing nach der Beantwortung eines jeden einzelnen Items. Der Vorteil des branched testing ist, dass auf den Computereinsatz verzichtet werden kann, denn es sind wesentlich weniger Entscheidungsschritte notwendig, die vom Testleiter live durchgeführt werden können (Kubinger & Jäger 2003, S. 5).

Das folgende Schema soll das Verfahren des branched testing verdeutlichen:

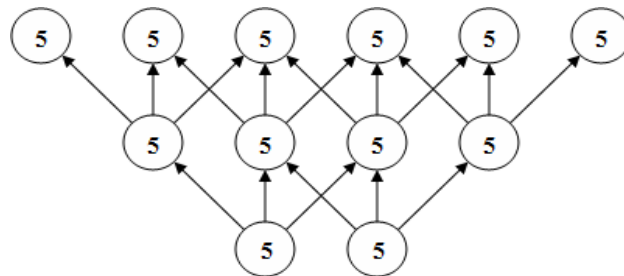


Abbildung 1: Beispiel eines Verzweigungsschemas beim branched testing

Jede Itemgruppe des dargestellten Beispiels enthält fünf Items. Der Proband wird aufgrund von Vorinformationen über die Ausprägung des zu untersuchenden Merkmals einer der beiden Startgruppen (unten) zugeordnet. Löst der Proband kein oder nur ein Item der gesamten Itemgruppe richtig, so wird ihm im nächsten Schritt eine leichtere Itemgruppe (weiter links) zugewiesen, löst der Proband vier oder fünf Items richtig, wird ihm im Folgenden eine Gruppe zugewiesen, deren Items einen höheren Schwierigkeitsgrad (weiter rechts) aufweisen als die bisher gelösten. Kann ein Proband etwa 50 Prozent der Items richtig beantworten, das wären in diesem Beispiel zwei bis drei, bleibt das Schwierigkeitsniveau in der nächsten zu lösenden Itemgruppe gleich. Im gezeigten Beispiel gibt es in der Testendphase sechs Gruppen unterschiedlicher Schwierigkeitsniveaus, die gemessene Leistung ergibt sich aus dieser Gruppe sowie aus der gelösten Itemanzahl innerhalb der Gruppe.

Das Ende des Tests wird ebenfalls durch einen Algorithmus bestimmt. Man unterscheidet *Zielkriterien* und *Abbruchkriterien* zum Beenden des Tests. Abbruchkriterien können das Ausschöpfen aller verfügbaren Items oder ein Erreichen der maximalen Testzeit sein. Ein Zielkriterium ist dann gege-

ben, wenn die Schwankung des Standardfehlers bei der Parameterschätzung minimiert ist und weitere Items keine weiteren Informationen zum zu untersuchenden Merkmals ergeben (Pospischill 2010, S. 155f.).

Einen Überblick über den Verlauf eines adaptiven Tests soll zusammenfassend die nachstehende Abbildung geben:

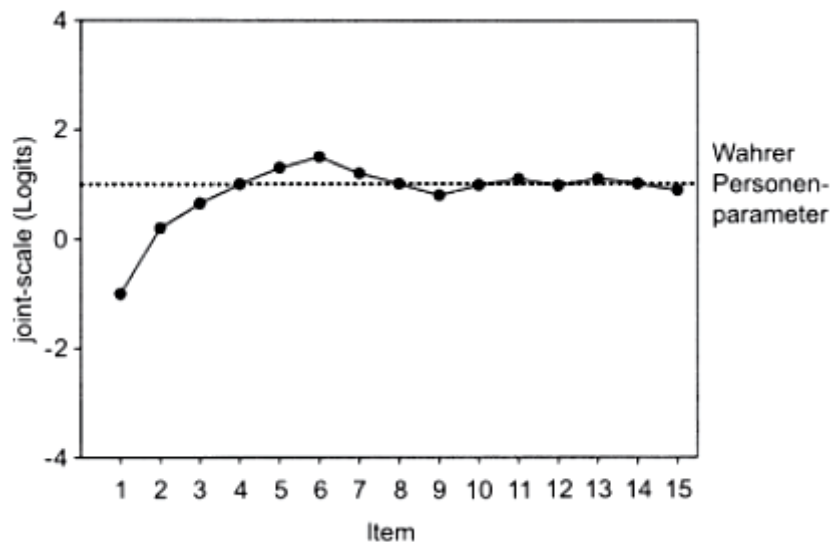


Abbildung 2: Illustration eines adaptiven Testablaufes (Frey 2008, S. 265)

Der Wert Null der Logitskala soll für eine mittlere Merkmalsausprägung stehen. Zu Beginn des in der Abbildung gezeigten Testes wurde ein Startitem mit einem niedrigen Schwierigkeitsgrad gewählt. Der Proband kann das erste Item richtig lösen, ebenso die folgenden vier Items. Er versagt erst beim sechsten Item und bekommt infolgedessen als siebentes Item ein solches mit einer geringeren Schwierigkeit. Item sieben und acht werden nicht gelöst, erst Item neun wird wieder richtig beantwortet. Dieses liegt leicht unter seinem wahren Personenparameter, bereits das vierte Item entsprach dem wahren Personenparameter des Probanden. Nach weiteren sechs Items kommt es zum Abbruch des adaptiven Testes, da weitere Items keine neuen Informationen bezüglich des Parameters aufzeigen können, da sich der Standardfehler der Schätzung des Personenparameters minimierte.

1.2 Vor- und Nachteile adaptiven Testens

Lehmann (2003) stellt eine Zusammenfassung der Vor- und Nachteile computergestützter Diagnostik auf (s. Tabelle 1). Adaptive Testverfahren sind – wie bereits erwähnt – häufig computerbasiert und aus diesem Grund sind ohne Einschränkung die Vor- und Nachteile der computergestützten Diagnostik auf adaptive computergestützte Testverfahren übertragbar.

Vorteile	Nachteile
<ul style="list-style-type: none"> • gute Akzeptanz • Messfehlerreduktion • Durchführung weniger anfällig für Fehler • zufällige Itemvorgabe und -reihenfolge möglich • hohe Standardisierung • schnellere, fehlerfreiere und erweiterte Auswertung • Verbesserung der diagnostischen Urteilsbildung • Messergebnisse können problemlos in Datenbanken aufgenommen werden • schnelles Feedback an Anwender • geringere zu bearbeitende Itemanzahl • höhere Motivation • Erfassung zusätzlicher Daten (Entscheidungszeit und Korrekturen) • Umweltfreundlichkeit • Entlastung des Personalarbeiters 	<ul style="list-style-type: none"> • kosten- und zeitintensive Neunormierung • hoher Entwicklungsaufwand • keine standardisierte Methodologie • Testleistung nicht unabhängig von Computerefähigkeiten (z.B. Tippgeschwindigkeit, Umgang mit der Computermouse) • eventuelle Computerangst: höhere Relevanz von Geschlecht und Alter

Tabelle 1: Vor- und Nachteile des computergestützten Testens (nach: Lehmann 2003, S. 45)

Um auch die Randbereiche eines Leistungsmerkmals differenziert erfassen zu können, empfiehlt es sich, für einen adaptiven Test einen großen Itempool bereit zu halten. Fisseni erwähnt als Untergrenze 100 Items (Fisseni 2004, S. 292). Weiterhin müssen diese Items homogen sein und „hinreichend das relevante Schwierigkeitskontinuum abdecken“. Stellen sich Items bei der Kalibrierung als inhomogen heraus, müssen diese ausgeschlossen werden. Decken die Items nicht alle Schwierigkeitsstufen ausreichend ab, kann es zu einer Über- oder Unterschätzung des zu messenden Merkmals kommen. Das Erstellen des geeigneten Itempools ist aufgrund des erhöhten Kosten- und Zeitaufwandes einer der größten Nachteile des adaptiven Testens (ebd., S. 291ff.).

Demgegenüber steht einer der größten Vorteile des adaptiven Testens: Die Durchführungs- und Auswertungsobjektivität sind deutlich höher als bei einem vergleichbaren Paper-Pencil-Test. Ebenso ist die Auswertung überhaupt nicht zeitaufwändig, da der Computer die Ergebnisse unmittelbar nach Testende ausgibt. Aus diesem Grunde, aber vor allem wegen der enormen Reduktion der Itemzahl und der damit verbundenen Testzeit kann von einer optimalen Testökonomie gesprochen werden.

Die kürzere Bearbeitungszeit des gesamten Testes kann auf den Probanden motivierend wirken, eventuell wird dieser Vorteil durch auftretende Computerangst eingeschränkt. Der nicht vertraute Umgang mit dem Computer kann – besonders bei älteren Probanden – zu Frustration, Demotivierung und damit zu einem verfälschten Testergebnis führen (ebd. S. 291ff.).

Adaptive Tests beginnen in der Regel mit einem Item mittlerer Schwierigkeit. Dies kann bei Probanden, deren Ausprägung des zu messenden Merkmals eher im unteren Bereich anzusiedeln ist, zu Überforderungen und Frustration führen. Umgekehrt kann ein mittelschweres Item von leistungsstarken Probanden als zu leicht empfunden werden und Demotivierung erzeugen.

Obwohl die probabilistische Testtheorie davon ausgeht, dass die Lösungswahrscheinlichkeit eines Items unabhängig von seiner Position in der Reihenfolge des Testes ist, widerlegen einige empirische Befunde diese Annahme. So ist beim adaptiven Testen nicht auszuschließen, dass Lernprozesse die Beantwortung weiterer Items beeinflussen.

Ein weiterer wichtiger Vorteil des computergestützten adaptiven Testens ist die Möglichkeit, zusätzliche Daten zu erfassen, wie zum Beispiel die benötigte Bearbeitungszeit der einzelnen Items oder eventuelle Korrekturen der Antworten.

Als Nachteil sehen Jäger & Krieger (1994) des Weiteren an, dass (zumindest zum Zeitpunkt der Publikation des Beitrags) keine Software auf dem Markt ist, die dem Interessenten am Kreieren eines eigenen adaptiven Tests eine benutzerfreundliche Struktur offeriert.

2. Das VBA-Projekt „Herstellung, Durchführung und Auswertung eines adaptiven Leistungstests“

2.1 Die Datei „adaptiv.xls“ als Gestaltungsrahmen

Mit dem Ziel, das adaptive Testen in der Lehrtätigkeit verständnisintensiver vermitteln zu können, entwickelte der Autor ein VBA-Projekt⁴, das einerseits anhand des geladenen Tests in passiver Weise das Wesen des Testverfahrens erkennen lässt und andererseits durch die Gestaltung, die Eichung und den Einsatz eines eigenen Tests ein noch tieferes Eindringen ermöglicht.

Das Projekt besteht aus folgenden Excelblättern, die vom Anwender mit eigenem Inhalt gefüllt und eingesetzt werden können:

herstellen

Hier werden der zu vergebende Dateiname, der Testtitel, die Instruktion für die Eichstichprobe

⁴ programmiert mit Visual Basic for Applications (Excel)

sowie für die Probanden des fertigen Tests, die Fragen mit ihren Auswahlantworten eingetragen. Das Blatt enthält vier Buttons für die Weiterarbeit:

1. der Button KONTROLLE, der die Vollständigkeit und formale Richtigkeit des Einzutragenden prüft,
2. der Button SORTIEREN, mit dem die Rangordnung des Schwierigkeitsgrades der Items nach dem Bearbeiten der Items durch die Eichstichprobe festgelegt wird,
3. der Button SPEICHERN, der die Ausgangsdatei *adaptiv.xls* unter dem selbst vergebenen Namen speichert, sowie
4. der Button ZUM TEST, der dem Probanden den fertigen Test vorlegt.

test_eichen

Hier wird das Mitglied der Eichstichprobe aufgefordert, seinen Namen einzugeben und es erhält nach dem Durchlesen der Instruktion konsekutiv in Zufallsfolge die Items mit den dazu gehörigen Auswahlantworten.

test

Der Aufbau dieses Blattes ist dem vorher beschriebenen ähnlich. Der Unterschied besteht darin, dass lediglich die ‚maßgeschneiderten‘ Items vorgelegt werden.

ergebnis

Dieses Blatt zeigt nach Beendigung der Arbeit am letzten Item den Namen des Probanden, eine Liste mit den Aufgabennummern, den Aufgabentexten, den abgegebenen Lösungen, den richtigen Lösungen, dem Vermerk ‚richtig‘ bzw. ‚falsch‘, die Summe der richtigen Antworten sowie die adaptive Testpyramide mit dem Pfad des Probanden, der den Testverlauf grafisch deutlich macht.

datenbank

Das Blatt ist vorrangig denjenigen Personen gewidmet, die sowohl beim Eichen des Tests mitarbeiteten als auch wesentlich später den fertigen Test absolvierten. Es zeigt einen Vergleich beider Leistungen bezüglich der adaptiv ausgewählten Testitems. Jedem Probanden wird eine Spalte zugeordnet. Ersichtlich ist der Lauf durch die Testaufgaben mit der Angabe der Richtigkeit der Lösung.

Mit grüner Farbe werden diejenigen Items hinterlegt, die die Person im Eichtest als auch im Echttest mit dem gleichen Prädikat bearbeitete (also entweder 2mal richtig oder 2mal falsch), dagegen werden diejenigen Items rot hinterlegt, wo es Abweichungen in der Richtigkeit der Lösung zwi-

schen Echttest und Eichprozess gab. Interessant sind dabei rote Zellen mit dem Prädikat richtig, denn hier könnte auch ein Erinnerungseffekt zum Tragen gekommen sein.

Die Spalten der Probanden, die zwar den Test bearbeiteten, aber nicht am Eichprozess teilnahmen, bleiben ohne Füllfarbe.

pfad

Im Unterschied zum Pfaddiagramm im Blatt ergebnis gestattet das Blatt Pfad die Darstellung der Pfade aller in der Datenbank enthaltenen Probanden. Dazu ist der Name des Probanden einzugeben (vgl. Abbildung 4 auf S. 16).

2.2 Der vorgefertigte Test „kopfrechnen_bvj_reha.xls“

2.2.1 Die Testitems

Dieser Test wurde mit dem Ziel einer Reliabilitätsuntersuchung hergestellt, nämlich nachzuweisen, dass er auf der Basis weniger „maßgeschneiderter“ Items zu einem verlässlichen Ergebnis führt. Die Erbringung dieser Aufgabe oblag der Mitautorin dieses Artikels Anita Polet im Rahmen ihrer Qualifikationsarbeit zur Magistra Artium der Erziehungswissenschaft.⁵

Für die Erstellung des Tests wurde ein Itempool angelegt, der aus 66 Kopfrechenaufgaben der vier Grundrechenarten (Addition, Subtraktion, Multiplikation und Division) besteht. Die Auswahl der Aufgaben erfolgte mit Unterstützung einer Mathematiklehrerin des Staatlichen Berufsbildenden Schulzentrums Jena-Göschwitz, Abteilung 3 (Lernbehinderung und Rehabilitation). Sie unterrichtet selbst in BVJ- und Rehaklassen und konnte somit das Anforderungsniveau der Berufsschüler der Eichstichprobe sehr gut einschätzen.

Innerhalb jeder Aufgabenart wurde darauf geachtet, ein möglichst breites Spektrum des Anforderungsniveaus abzudecken, um auch die Leistungen in den Randbereichen differenziert erfassen zu können.

Die Aufgaben der Addition bestehen dabei aus zwei bis drei Summanden⁶ und umfassen sowohl natürliche als auch rationale Zahlen⁷. Bei der Auswahl der Aufgaben der Addition wurde darauf geachtet, dass unterschiedlich schwere Aufgaben diesem Komplex angehören, so wird die Kenntnis des Zehner- und Hunderterystems und das damit verbundene Kopfrechnen mit Zehner- und Hunderterübergängen überprüft⁸. Die Aufgaben der Subtraktion wurden nach den gleichen Kriterien wie die Aufgaben der Addition ausgewählt. Die Multiplikationsaufgaben umfassen neben Auf-

⁵ Polet, A.: Eine Reliabilitätsuntersuchung zum adaptiven Testen mit der Testpyramide anhand der Kompetenz in Grundrechenarten bei lernschwachen Berufsschülern.

⁶ zum Beispiel: $236 + 236$ (zwei Summanden), Aufgabe $243 + 57 + 183$ (drei Summanden)

⁷ zum Beispiel: $324 + 293$ (natürliche Zahlen), und Aufgabe $4,5 + 5,5$ (rationale Zahlen)

⁸ zum Beispiel: $419 + 224$ (Zehnerübergang) und Aufgabe $190 + 1352$ (Hunderterübergang)

gaben des kleinen Einmaleins⁹ auch Aufgaben mit je einem Faktor der natürlichen und rationalen Zahlen¹⁰, sowie Aufgaben, die über das kleine Einmaleins hinausgehen¹¹. Die Aufgaben der Division umfassen Aufgaben des kleinen Einmaleins¹², Aufgaben mit ‚glatte‘ Zehnern¹³ oder dem Term 2^{14} . Einige Aufgaben gehören zum Aufgabenkomplex der Mischformen¹⁵, da dieser voraussetzt, dass sowohl die Grundrechenarten, als auch die Regel Punkt- vor Strichrechnung, beherrscht werden müssen. Zwei weitere Aufgaben weichen vom Prinzip der Grundrechenarten ab und verlangen die Fähigkeit des Transkodierens, also das Übertragen der Zahlen aus der Wortform in die Zahlenform.

Die Tabelle 2 ist ein Ausschnitt aus dem Tabellenblatt herstellen. Sie umfasst alle 66 Aufgaben mit ihren Auswahlantworten.

Nr.	Aufgabe	Antwort 1	Antwort 2	Antwort 3	Antwort 4	...
1	$17000 + 48 =$	17480	17408	17480	17048+	
2	$4,5 + 5,5 =$	9	10+	9,10	10,10	
3	$48 : 6 =$	6	8+	7	5	
4	$10 * 24 =$	24	240+	2400	24000	
5	$500 - 0,5 =$	499,5+	500,5	250	10	
6	$760 - 111 =$	649+	651	871	671	
7	$488 - 303 =$	153	185+	785	753	
8	$6,8 + 7 =$	13,8+	12,8	14,8	15,8	
9	$22000 - 5 =$	22005	21955	21995+	21500	
10	$1,2 * 3 =$	5,6	4	6	3,6+	
11	$9 * 0 =$	0,9	9	0+	1	
12	$56 : 7 =$	7	6	9	8+	
13	$80 * 50 =$	400	4000+	40	130	
14	$7 + 3 * 5 =$	50	22+	8	21	
15	Bilde die Summe aus 9 und 3	27	12+	3	6	
16	$420 : 6 =$	7	60	6	70+	
17	$28 : 4 =$	9	7+	6	8	
18	$1 * 0 * 2 =$	0+	2	1	3	
19	$18,9 + 6,1 =$	24,8	25+	26	25,8	
20	$23 + 17 + 38 =$	72	71	75	78+	
21	Addiere 8 und 5	40	14	12	13+	
22	$5 * 5 * 5 =$	100	125+	15	50	
23	$15 * 6 =$	90+	111	130	75	
24	$4 * 2 * 9 =$	56	15	81	72+	
25	$16 * 3 =$	42	36	32	48+	
26	$100 : 4 =$	15	25+	20	30	
27	Bilde das Produkt aus 9 und 3	12	3	6	27+	
28	$500 : 50 =$	50	10+	5	100	

⁹ zum Beispiel: Bilde das Produkt aus 9 und 3

¹⁰ zum Beispiel: $0,5 * 6$

¹¹ zum Beispiel: $25 * 9$

¹² zum Beispiel: $56 : 7$

¹³ zum Beispiel: $500 : 50$

¹⁴ zum Beispiel: $982 : 2$

¹⁵ zum Beispiel: $108 - 8 * 8$

29	$56,3 - 0,8 =$	55,1	56,1	55,1	55,5+
30	$4 * 8 + 120 =$	142	152+	512	412
31	$52 : 26 =$	2+	3	4	0,5
32	$324 + 293 =$	617+	611	627	531
33	$419 + 224 =$	643+	633	605	635
34	$25 * 9 =$	325	1845	225+	350
35	$600 * 90 =$	6300	56000	54000+	4800
36	$60 : 12 =$	5+	7	0,6	6
37	$700 - 64 =$	654	646	636+	764
38	$243 + 57 + 183 =$	383	393	493	483+
39	$14 - 8 : 2 =$	10+	3	4	8
40	$910 - 750 =$	1660	240	160+	140
41	$75 * 3 =$	175	375	450	225+
42	$0,5 * 6 =$	3+	30	0,3	0,30
43	Bilde die Hälfte von 728	314	364+	414	464
44	$1467 - 69 =$	1396	1402	1400	1398+
45	Bilde das Doppelte von 368	626	736+	762	726
46	$56,3 - 6,8 =$	55,1	50	50,5	49,5+
47	$242 - 119 =$	121	123+	141	133
48	$7 * 9 + 241 =$	1650	402	1750	304+
49	$190 + 1352 =$	1442	1542+	1552	1452
50	$982 : 2 =$	441	1964	491+	451
51	$306 : 2 =$	153+	103	612	149
52	$680 - 34 - 47 =$	591	601	589	599+
53	$45,6 + 163,4 =$	209+	199	209,2	219,2
54	Zwölfmillionensiebenundvierzigtausendacht	12047008+	124708	12478	1204708
55	Bilde den Quotienten aus 9 und 3	3+	12	27	6
56	Bilde die Differenz aus 9 und 3	6+	27	3	12
57	$123,8 + 42,6 =$	165,14	166,2	166,4+	165,4
58	$260 + 280 - 190 =$	350+	450	550	250
59	$1,6 * 5 =$	8,1	5,9	8+	5,1
60	$1366 - 229 =$	1127	1143	1133	1137+
61	$108 - 8 * 8 =$	800	100	64	44+
62	$2986 + 842 =$	3728	3828+	3848	3738
63	$7 * 4 + 3 * 9 =$	45	441	55+	61
64	$24,6 - 19,7 =$	4,9+	5,3	4,3	5,9
65	$55 + 6 * 8 =$	420	110	488	103+
66	Sechsmillioneneinhundertzweiundneunzigtausenddrei- undsechzig	6192063+	619263	60192063	6192630

Tabelle 2: Die Aufgaben und ihre Auswahlantworten

Die Anzahl der Aufgaben des Eichtestes wurde auf 66 begrenzt, denn es sollte für die an der Eichung des Tests¹⁶ beteiligten Schüler möglich sein, inklusive der Vor- und Nachbereitungstätigkeiten alle Aufgaben des Tests innerhalb einer Unterrichtsstunde zu lösen.

Die 66 Testaufgaben sind nach dem Single-Choice-Verfahren zu lösen. Für jedes Testitem wurden 4 Antwortmöglichkeiten kreiert, wobei bei der Gestaltung der falschen Antworten Absurdes nicht

¹⁶ Die zur Testeichung herangezogenen Probanden müssen alle Aufgaben lösen – im Gegensatz zu den Probanden, deren Leistung später mit dem Test gemessen wird.

zugelassen wurde sondern die von der o.g. Lehrerin benannten typischen Fehler berücksichtigt wurden (vgl. dazu auch Lienert & Raatz 1993, S. 19). Dies soll eine der Aufgaben verdeutlichen:

Aufgabe 10: $4,5 + 5,5$			
a 9	b 10	c 9,10	d 10,10

Alle vier Antwortmöglichkeiten sind denkbar. Kann der Proband nicht mit rationalen Zahlen umgehen, so wäre es möglich, dass die Nachkommastellen beim Kopfrechnen weggelassen werden (Antwort a), das Komma als Trennung aufgefasst wird und sowohl die Stellen vor als auch nach dem Komma einzeln und unabhängig voneinander addiert werden (Antwort c), die Kenntnis über das Addieren von rationalen Zahlen teilweise vorhanden ist und die Zahl vor dem Komma richtig berechnet wird, jedoch nicht erkannt wird, dass nicht immer beim Addieren rationaler Summanden auch die Summe eine rationale Zahl sein muss (Antwort d). Richtig ist für das Beispiel die Antwort b.

Um Zufallslösungen zu vermeiden, war es wichtig, dass die richtige Lösung nicht immer die gleiche oder eine systematisch veränderte Position in der Reihenfolge der Antwortmöglichkeiten einnahm sondern bei jeder Aufgabe zufällig platziert wurde.

Um Effekte bezüglich der Reihenfolge der Aufgaben aufeinander auszuschließen, enthielt das Programmscript einen Passus, der alle 66 Aufgaben in einer Zufallsreihenfolge vorlegt.

Der in unserer Arbeit verwendete adaptive Testmodus ist der der Testpyramide, die das oben beschriebene ‚tailored testing‘ realisiert. Unsere Pyramide besteht aus zehn Ebenen: Ebene 1 besteht dabei aus einer Aufgabe, Ebene 2 besteht aus zwei Aufgaben, usw. Die letzte Ebene besteht demzufolge aus 11 Aufgaben und damit der gesamte Itempool aus $1+2+3+4+5+6+7+8+9+10+11=66$ Aufgaben (vgl. Abbildung 3).

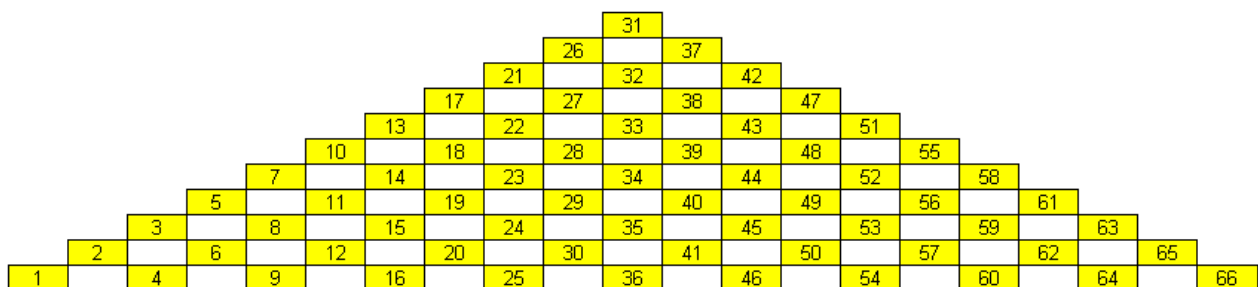


Abbildung 3: Struktur der adaptiven Testpyramide mit 11 Ebenen

2.2.2 Die Testeichung

Mit dem Ziel der Erzielung eines möglichst validen Eichwertes wurde von der Eichstichprobe neben der Richtigkeit der Lösung der Aufgabe auch die benötigte Zeit für die Lösung der Aufgabe festgehalten. Beide Variablen bestimmen den Eichwert nach der nachstehenden Formel:

$$\text{Eichwert} = \frac{i+}{\bar{x}_{t+}}$$

$i+$ Anzahl der richtigen Lösungen

\bar{x}_{t+} Mittelwert der benötigten Zeit in Sekunden der Richtiglöser

Dieser Eichwert lässt schließlich das Ranking der einzelnen Aufgaben und damit die in Abbildung 3 gezeigte Struktur ermitteln.

Es kommt häufig vor, dass Aufgaben von gleich vielen Probanden richtig gelöst werden. In diesem Fall entscheidet die benötigte Zeit über den Eichwert und damit über die Rangposition. Im ganz besonderen Fall der Gleichheit zweier Eichwerte zeigt das Programm nur einen Rangplatz an und fordert dazu auf, entweder weitere Probanden in den Eichprozess einzubeziehen oder – falls dies aus organisatorischen Gründen nicht möglich ist – eine auf Erfahrung beruhende Teilrangfolge – zu bestimmen.¹⁷ Zusammenfassend lässt sich festhalten: Je mehr Probanden eine Aufgabe richtig lösen und je weniger Zeit sie für die richtige Lösung im Durchschnitt benötigen, desto leichter wird diese Aufgabe gerankt – und damit in der Struktur weiter links stehend. Im Gegensatz dazu ist eine Aufgabe schwer, die von nur wenigen Probanden richtig gelöst wird und dafür eine längere Lösungszeit im Mittel erfordert – in der Struktur weiter rechts stehend.

Aus der Rankingnummer (siehe Tabelle 3) ergibt sich die Position der Aufgabe in der Testpyramide: 1 ist die leichteste und 66 die schwerste Aufgabe, Aufgabe 31 ist gleich dem Median. Aus der Testpyramide könnte man entnehmen, dass untereinander stehende Aufgaben gleich schwer sein müssten. Dies ist jedoch nicht der Fall: Die Eichwerte untereinander stehender Aufgaben differieren leicht. Eingehalten ist jedoch durchweg, dass ein aus der Struktur beliebig herausgegriffenes Dreieck oben den mittleren Schwierigkeitsgrad, darunter links den geringsten und darunter rechts den höchsten innerhalb der drei besitzt. Dieser Schluss ergibt sich aus dem Aufbau der Pyramide: Sie setzen sich von links betrachtet von 1 bis 66 in jeder Spalte von oben nach unten fort.

¹⁷ Zu beachten ist, dass die Eichwerte durchaus nicht gleich sein müssen, weil das Programm auf 2 Nachkommastellen gerundete Werte anzeigt. Eine Lücke wird nur in ganz seltenen Fällen entstehen.

Die Tabelle 3 zeigt einen weiteren Ausschnitt aus dem Blatt herstellen. Ersichtlich sind wiederum die Aufgaben, der Eichwert, der Rang sowie die jeweils ersten drei Ergebnisspalten und Zeitspalten.¹⁸

Nr.	Formulierung der Frage	...	Eichwert	Rang	Erg S1	Erg S2	Erg S3	...	Time S1	Time S2	Time S3
1	17000 + 48 =		8,61	1	1	0	0		4,83	5,39	8,70
2	4,5 + 5,5 =		7,07	2	0	1	0		8,86	3,64	24,25
3	48 : 6 =		6,58	3	1	1	0		5,30	2,97	0,52
4	10 * 24 =		5,66	4	1	1	0		4,13	3,11	3,31
5	500 - 0,5 =		5,63	5	1	1	1		8,64	15,59	7,17
6	760 - 111 =		5,62	6	1	1	1		6,92	17,70	8,73
7	488 - 303 =		5,58	7	1	1	0		9,75	12,29	0,19
8	6,8 + 7 =		5,51	8	0	1	0		16,06	5,13	39,52
9	22000 - 5 =		5,10	9	0	1	1		8,60	11,80	6,11
10	1,2 * 3 =		5,08	10	1	1	0		9,50	29,18	13,53
11	9 * 0 =		5,07	11	1	0	1		5,11	26,01	5,19
12	56 : 7 =		4,95	12	1	1	0		4,77	5,55	4,39
13	80 * 50 =		4,93	13	0	1	0		11,06	11,75	4,64
14	7 + 3 * 5 =		4,86	14	1	1	0		5,21	4,47	0,42
15	Bilde die Summe aus 9 und 3		4,79	15	1	0	0		4,32	5,00	0,18
16	420 : 6 =		4,72	16	1	1	0		27,62	13,39	12,33
17	28 : 4 =		4,60	17	0	1	0		6,20	11,63	10,16
18	1 * 0 * 2 =		4,53	18	1	0	0		4,20	5,30	37,75
19	18,9 + 6,1 =		4,22	19	0	1	0		13,76	8,56	4,13
20	23 + 17 + 38 =		4,08	20	1	1	1		8,72	18,50	7,70
21	Addiere 8 und 5		4,01	21	0	1	1		6,09	6,82	8,59
22	5 * 5 * 5 =		4,00	22	1	1	0		4,47	5,42	0,19
23	15 * 6 =		3,76	23	0	1	1		4,53	7,95	60,71
24	4 * 2 * 9 =		3,57	24	0	1	0		8,72	18,82	4,58
25	16 * 3 =		3,46	25	0	1	0		14,74	21,23	1,02
26	100 : 4 =		3,46	26	0	1	0		24,50	6,14	4,55
27	Bilde das Produkt aus 9 und 3		3,40	27	0	1	0		4,56	6,92	34,17
28	500 : 50 =		3,29	28	0	1	0		19,99	3,95	19,33
29	56,3 - 0,8 =		3,20	29	1	0	0		20,74	16,51	2,39
30	4 * 8 + 120 =		3,20	30	0	1	0		27,28	15,48	40,50
31	52 : 26 =		3,02	31	0	0	0		8,04	2,28	10,03
32	324 + 293 =		2,97	32	1	1	1		13,78	25,55	33,23
33	419 + 224 =		2,93	33	0	1	0		8,95	39,87	31,98
34	25 * 9 =		2,88	34	1	1	0		5,30	12,11	37,83
35	600 * 90 =		2,63	35	0	1	1		13,38	2,23	8,78
36	60 : 12 =		2,62	36	0	1	0		11,03	28,07	48,42
37	700 - 64 =		2,50	37	0	0	0		8,68	8,67	9,40
38	243 + 57 + 183 =		2,49	38	1	0	0		29,87	10,20	9,47
39	14 - 8 : 2 =		2,39	39	0	1	0		11,77	9,72	0,20
40	910 - 750 =		2,38	40	0	1	0		28,86	12,82	10,08
41	75 * 3 =		2,29	41	0	1	0		11,95	12,63	16,78
42	0,5 * 6 =		2,28	42	0	1	0		36,15	91,04	22,14
43	Bilde die Hälfte von 728		2,25	43	1	1	0		31,11	18,54	20,67
44	1467 - 69 =		2,23	44	1	1	0		13,27	76,02	16,69
45	Bilde das Doppelte von 368		2,18	45	0	1	0		19,16	31,12	29,68
46	56,3 - 6,8 =		2,14	46	0	0	0		5,29	32,37	8,89

¹⁸ Ergebnis: 0 = falsch gelöst, 1 = gelöst / Zeit: in Sekunden

47	242 - 119 =	2,12	47	0	1	1	23,84	43,33	37,39
48	7 * 9 + 241 =	2,12	48	0	1	0	4,95	11,29	148,85
49	190 + 1352 =	1,93	49	0	1	0	15,61	55,24	6,45
50	982 : 2 =	1,77	50	0	1	1	4,34	15,12	10,77
51	306 : 2 =	1,70	51	0	1	0	16,82	12,28	22,19
52	680 - 34 - 47 =	1,68	52	0	0	0	14,19	12,01	3,72
53	45,6 + 163,4 =	1,62	53	1	0	0	13,09	14,00	2,17
54	Zwölfmillionensiebenundvierzigtausend-acht	1,59	54	1	1	0	17,20	17,84	33,39
55	Bilde den Quotienten aus 9 und 3	1,59	55	1	1	0	10,89	1,95	21,70
56	Bilde die Differenz aus 9 und 3	1,54	56	1	0	0	18,23	24,54	10,28
57	123,8 + 42,6 =	1,53	57	0	0	1	7,42	21,54	1,67
58	260 + 280 - 190 =	1,53	58	1	1	0	19,26	28,79	11,40
59	1,6 * 5 =	1,39	59	0	1	0	9,76	25,25	14,50
60	1366 - 229 =	1,35	60	0	0	1	15,96	1,59	0,29
61	108 - 8 * 8 =	1,35	61	0	0	0	10,52	2,93	2,95
62	2986 + 842 =	1,35	62	0	0	0	17,46	43,76	19,48
63	7 * 4 + 3 * 9 =	1,15	63	0	1	0	25,73	28,16	8,06
64	24,6 - 19,7 =	1,15	64	1	0	0	36,96	24,20	19,28
65	55 + 6 * 8 =	1,10	65	0	1	0	24,25	26,29	16,17
66	Sechsmillioneneinhundertzweiundneunzigtausenddreihundsechzig	1,09	66	1	1	0	52,74	37,29	44,13

Tabelle 3: Aufgaben, Eichwerte, Ranking, Ergebnisse und Zeiten

(Ausschnitt: Aus Platzgründen konnten hier nur die Ergebnisse und die Zeiten der drei ersten Probanden in die Tabelle aufgenommen werden.)

Damit ist der Eichungsprozess abgeschlossen und der adaptive Test kann nunmehr eingesetzt werden.

2.2.3 Der Testeinsatz

Die erste Aufgabe, die im Echtttest jeder Proband bekommt, ist die Aufgabe mit mittlerem Schwierigkeitsniveau (Rankingnummer 31). In Abhängigkeit von der Antwort des Probanden erfolgt die Auswahl der zweiten Aufgabe. Wird die Aufgabe 31 richtig beantwortet, bekommt der Proband die nächst schwierigere Aufgabe (Ranking 37). Wird die erste Aufgabe hingegen falsch gelöst, so erhält er die nächst leichtere Aufgabe (Ranking 26). Dieses Verfahren wird bis zur zehnten Ebene fortgesetzt, so dass der Proband anstelle der 66 Items nur mit 11 Items konfrontiert wird und damit 55 Aufgaben (das sind ca. 83% der Itemmenge) einspart.

Während der Ausgangspunkt für alle Testprobanden gleich ist, kann ihr Endpunkt in der Ebene 11 von Aufgabe 1 (d.h. keine der Aufgaben richtig gelöst) bis zur Aufgabe 66 (d.h. alle Aufgaben richtig gelöst) rangieren. Im Allgemeinen stellt sich jedoch nicht einer dieser Extremfälle ein sondern ein Zickzackpfad. Die Leistung des Probanden wird in unserer Arbeit also auf eine 11-stufige Skala (0 ... 10 Punkte) abgebildet. Die Abbildung 4 zeigt den Pfad durch die Testpyramide für einen ausgewählten Probanden:



Abbildung 4: Screenshot des Excelblattes pfa

Der Leser dieses Artikels möge beachten, dass die beschriebene Arbeit in erster Linie zu Lehrzwecken entstand. Wenngleich die Mitautorin Anita Polet in ihrer Magisterarbeit zeigen konnte, dass im Rahmen der Ermittlung der Test-Retest-Reliabilität eine ausreichende Korrelation zwischen den Variablen Ergebnis im Eichtest und Ergebnis im adaptiven Test berechnet werden konnte ($N=57$; $r_{tt}=.669$), sollte die Arbeit nur bedingt zur Herstellung ‚echter Tests‘ verwendet werden. Bei gleichem Verfahren sollten hierfür mehr als 11 Strukturebenen einbezogen werden.

Sollte dieser Beitrag Ihr Interesse geweckt haben, dann nehmen Sie bitte Kontakt mit Dr. Helmut Stauche (stauche46@gmail.com) auf. Sie erhalten dann die Dateien **adaptiv.xls** (leere Datei zum Gestalten eines eigenen adaptiven Tests) sowie die Datei **kopfrechnen_bvj_reha.xls** (unveränderbare, mit Inhalt gefüllte Datei zum Ausprobieren und Vorführen).

Nach dem Einsatz in der Lehre bitten wir Sie um ein Feedback zum evtl. Erfolg bzw. zu Problemen beim Einsatz der Programme.

Referenzen

- Fisseni, H.-J. (³2004): Lehrbuch der psychologischen Diagnostik. mit Hinweisen zur Intervention. Hogrefe: Göttingen.
- Frey, A. (2008): Adaptives Testen. In: Moosbrugger, H. & Kelava, A.: Testtheorie und Fragebogenkonstruktion. Heidelberg: Springer, S. 260-278.
- Kubinger, K. D. & Jäger R. S. (Hrsg.) (2003): Schlüsselbegriffe der psychologischen Diagnostik. Weinheim: Beltz.
- Jäger R. S. & Krieger, W. (1994): Zukunftsperspektiven der computergestützten Diagnostik, dargestellt am Beispiel der treatmentorientierten Diagnostik. Diagnostica 40, 217-243
- Lehmann, K. (2003): Auswahl von Mitgliedern virtueller Teams. Entwicklung und Validierung eines Online-Testverfahrens. Wiesbaden: DUV.
- Lienert, G. A. & Raatz, U. (⁵1993): Testaufbau und Testanalyse. Weinheim: Beltz.
- Polet, A. (2011): Eine Reliabilitätsuntersuchung zum adaptiven Testen mit der Testpyramide anhand der Kompetenz in Grundrechenarten bei lernschwachen Berufsschülern. Magisterarbeit zur Erlangung des akademischen Grades Magistra Artium. Jena: Universitätsarchiv.
- Pospeschill, M. (2010): Testtheorie, Testkonstruktion, Testevaluation. Mit 77 Tabellen zur Wiederholung. München: Reinhardt.